On the use of Jargon and Word Embeddings to Explore Subculture within the Reddit's Manosphere

Conference or Workshop Item

On the use of Jargon and Word Embeddings to Explore Subculture within the Reddit's Manosphere

Tracie Farrell
tracie.farrel@open.ac.uk
Open University

Oscar Araque
Universidad Politécnica de Madrid
o.araque@upm.es

Miriam Fernandez
Open University
miriam.fernandez@open.ac.uk

Harith Alani
Open University
h.alani@open.ac.uk

ABSTRACT

Understanding the identities, needs, realities and development of subcultures has been a long term target of sociology and cultural studies. Socio-cultural linguistics, in particular, examines the use of language and, in particular, the existence and use of neologisms, slang and jargon. These terms capture concepts and expressions that are not in common use and represent the new realities, norms and values of subcommunities. Identifying and understanding such terms, however, is a very complex task, particularly considering the vast amount of content that is currently available online for many such groups. In this paper, we propose a combination of computational and socio-linguistic methods to automatically extract new terminology from large amounts of data, using word-embeddings to semantically contextualise their meaning. As a use case, we explore subculture on the platform Reddit. More specifically, we investigate groups considered part of the manosphere, a loose online community where men's perspectives, gripes, frustrations and desires are explicitly expressed and where women are typically targets of hostility. Characterisations of this group as a subculture are then provided, based on an in-depth analysis of the identified jargon.

CCS CONCEPTS

• **Computing methodologies** → **Information extraction**; Topic modeling; • **Theory of computation** → *Random projections and metric embeddings.*

KEYWORDS

word embedding, jargon, semantics, manosphere, misogyny, Reddit

ACM Reference Format:
Tracie Farrell, Oscar Araque, Miriam Fernandez, and Harith Alani. 2020. On the use of Jargon and Word Embeddings to Explore Subculture within the Reddit's Manosphere. In *12th ACM Conference on Web Science (WebSci '20), July 6–10, 2020, Southampton, United Kingdom.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3394231.3397912

WebSci '20, July 6–10, 2020, Southampton, United Kingdom

ACM ISBN 978-1-4503-7989-2/20/07…$15.00
https://doi.org/10.1145/3394231.3397912

1 INTRODUCTION

Subcultures and subsocieties (hippies, goths, bikers, Harry Potter fans) are smaller groups within a larger society, which share values and norms that are sometimes distinct from those held by the majority [13].[1] Subcultures can have their own divisions and hierarchies within them [24], and frequently develop specialised vocabularies that reflect unique aspects of the identity, needs and realities of group members [27]. In the study described in this paper, we use the phenomenon of language innovation in the manosphere (see 3) to understand how different groups within it view their own departure from the mainstream society. Through the prism of specialised vocabulary, we examine some of the factors that may link groups in the manosphere to hate, crime and violence.

Socio-cultural linguistics addresses the structure and meaning of specialised vocabularies (neologisms, slang, jargon). Before computational approaches began to influence this work, investigations were typically based on in-depth observations of the subculture's rhetoric, and usually conducted over small sets of data. However, through the internet and social media, many sub-cultural groups have emerged, with vast amounts of content (discussions, interactions, etc.) now available. This content spans large periods of time. Manual identification and analysis of the subcultures' specialised vocabularies has become impractical and **automatic methods** that help with both the **identification** and **understanding** of specialised terms are needed.

In this work we propose an interdisciplinary approach for **investigating emergent vocabularies** (which we refer to as **jargon** [7]) of online subcultures. We address novel jargon, or "neologisms" in the context of a subculture. Our approach proposes a combination of: (i) **computational methods**, to automatically identify and semantically contextualised jargon terms with word-embeddings (validating understanding of the terms' meanings within the subculture) and, (ii) **socio-linguistic methods**, where a categorisation of the subculture is developed based on in-depth studies of the literature. The automatically identified jargon terms are then manually assessed, labelled and analysed based on those categories. The use case selected to showcase the application of our proposed approach is the **manosphere, a men's interests subculture online that has been linked to violent crime** (see section 3). Seven different Reddit communities have been selected as representative of this subculture or subsociety (see section 5).[2]

[1] https://haenfler.sites.grinnell.edu/subcultural-theory-and-theorists/what-is-a-subculture/
[2] Subsocieties may not always share values [13]

The data extracted from these communities includes 6 million posts, from 300K conversations created between 2011 and 2019.

The research questions and key contributions associated with this study can be summarised as:

- *RQ1: How can we automatically identify the specialised novel vocabulary (jargon) from a subcultures' generated textual content?* We propose a novel approach based on Natural Language Processing that identifies, filters and semantically contextualises a subculture's jargon, based on provided textual data. We have applied this approach to identify the manosphere's specialised vocabulary, extracting 2,615 terms.
- *RQ2: How can we combine socio-linguistic and computational methods for studying a subcultures' specialised vocabulary?* First, we conducted an in-depth study of the subculture and characterise it by means of ten thematic categories and four subcategories (e.g., opposition to feminism, conflict of masculinity). These categories have then been used to manually label each of the 2,615 identified jargon terms. Six 'Exclusion categories' have also emerged during the assessment of the extracted jargon that describe errors in the identification process (e.g., terms that refer to the online platform's jargon, Reddit, rather than the jargon of the subculture). Further analysis have then been conducted over the curated and labelled data as means to study the subculture's identify and interests.
- *RQ3: Can word-embeddings increase our ability to understand the meaning of jargon?* Understanding the meaning of the subculture's specialised vocabulary requires in-deep knowledge of the subculture. A key aspect of our proposed approach is the use of word-embeddings to contextualise jargon and facilitate its analysis, interpretation and semantic understanding. As a way to assess whether word-embeddings do indeed facilitate the understanding of jargon, the results of the human annotation have been compared against the generated word-embeddings by calculating the distance within the embeddings pace. Our results show that the word groupings identified by the human annotator do indeed align with the grouping of the terms in the embedding space.

The rest of the paper is structured as follows: Section 2.1 describes the manosphere, the subculture selected as use case in this work. Section 2 summarises related work from social, computational and interdisciplinary fields related to the identification and analysis of jargon, particularly in the manosphere. Section 3 is an in-depth analysis of the literature of the manosphere and the key characterisations of the subculture that emerge from previous studies. Section 4 describes our proposed automatic jargon extraction methodology. Sections 5 and 6 describe how jargon is identified, annotated, validated and analysed. Sections 7 and 8 discuss our results and conclude the work.

2 RELATED WORK

In this section we briefly present our use case of the manosphere (extended in 3). We describe some key Socio-linguistic and computational approaches that have attempted to identify and research jargon as a cultural and social object for different subcultures and the manosphere in particular.

2.1 Use Case: The Manosphere

While many subcultures and subsocieties are not harmful or offensive, others (such as neo-Nazi skinheads, football hooligans, or the manosphere[3]) often promote hate and have sometimes been linked with hate crimes, radicalisation, extremism and terror attacks.

The manosphere is a loose collection of online groups, in which members promote **"men's issues"** [20], such as father's rights, forced conscription and war, violence against men, homelessness and mental health, as well as attainment gaps in education and the workforce[4]. These groups often allow open hostility and misogyny toward women [20]. Groups that belong to the manosphere include some groups of men's rights activists, Men Going Their Own Way (MGTOW), pick-up artists (PUAs), and more recently, involuntarily celibates (Incels) [14, 30]. High profile crimes have been connected to manosphere communities[1, 4, 22] and have prompted the question of **whether or not these groups hold extremist views**.[5] If so, what do such groups offer their members? Studies of jargon provide insight into community development [27] and identification [21], through highlighting what is missing. Knowing which new concepts have emerged, around which major themes, will help us to understand more about what connects and differentiates communities in the manosphere.

2.2 Socio-Linguistic Approaches

Socio-linguistic approaches to identify and study jargon include language analysis of lexical innovation and the **function of jargon in the community**. Elaine Chaika, for example, [7] examined how syntactic and lexical features of specialist vocabulary used by two different communities could be analysed to reveal characteristics of the community or specific needs. She studied jargon used by bowlers and truck drivers speaking on the Citizen's Band radio (CB), focusing on the process of jargon innovation and language change. She found that CBers did not create lexical innovations to help devise a short-hand, but to entertain themselves without the need for efficiency (e.g."pregnant rollerskate" as a word for a Volkswagen). This was in contrast to bowlers who have short lexical innovations, like the word "Turkey" for scoring three strikes in a row. Chaika argues that language changes are not random or needless, but "extend old material to new situations when the need arises". Schulz also identified patterns of efficiency in jargon use, in an ethnographic study of medical jargon used in hospital libraries. Her participants needed efficient communication under conditions of time-pressure, in a highly specialised environment [36].

Studies of jargon among black communities in the states revealed additional purpose - to **communicate presence**, the need to **respond to hegemonic culture** and language use [8]. In fact, many minority or undervalued groups do identify jargon as a cultural symbol within their communities[27]. In our work, we are looking for both self- and external characterisations in the manosphere through analysing their jargon. By triangulating what is there in the data, with what we know from more in-depth studies of the

[3]We refer to the manosphere as a subculture as it has been defined elsewhere in Lilly *et al* [25] and Ging [15].
[4]http://content.time.com/time/business/article/0,8599,2015274,00.html
[5]http://www.nottinghamwomenscentre.com/wp-content/uploads/2018/07/Misogyny-Hate-Crime-Evaluation-Report-June-2018.pdf

manosphere, we can gain a better understanding of how this group relates to mainstream society.

Ging, for example, used an inductive approach, gathering data on the links between different men's groups online over a period of 6 months to develop a theoretical analysis of their network. She describes how jargon around alphas and betas in the manosphere represents a conflict of masculinity in which men struggle against one another for access to **power within the dominant group**[15]. Providing evidence of this, at least in part, Schmitz and Kazyak studied Men's Rights Activists (MRAs) on 12 popular websites using discourse analysis. Cataloguing their interactions, the authors identified two categories of members, **Cyber Lads in Search of Masculinity** and **Virtual Victims in Search of Equality** [35]. Whereas cyber lads tended to focus more on **degrading women and exercising aggression**, virtual victims tended to **adopt the rhetoric of social movements** to frame their experiences as those of an **undervalued or excluded community**. These features of the manosphere will be discussed in more detail in section 3. While valuable on it's own, this work is time-consuming and could be amplified. In our work, we are using these studies to inform our annotation process and guide our computational study, exploring and validating some of their claims at scale.

2.3 Computational Approaches

Computational approaches for studying jargon typically focus on vocabulary, and use natural language processing to explore patterns in language use across different data corpora. Wurschinger *et al* [40] tracked the use of competing synonymous neologisms on Twitter and the Web, using a Web-as-Corpus approach, to follow the progression of words through the network. Their analysis of Twitter provided evidence of neologisms at their early stages, while their Web studies illustrated the process of conventionalisation. The authors argue that social media platforms contribute to the spread of neologisms, and that **platform language conventions have an influence**.

Veale *et al* [39] used natural language processing techniques to identify and define blended words (like 'chunnel', a blend of the words 'channel' and 'tunnel') and portmanteaus (like 'incel', a combination of 'involuntary' and 'celibate'). This and work in this vein utilises knowledge about etymology and semantic changes to understand more about the syntax or semantics of neologisms. The authors include lexical blends, text messaging forms and ameliorations/pejorations (such as 'bad' for good or 'sick' for good)[10] in their study.

Other large bodies of research involve identifying the strength and emergence of neologisms for the purposes of advancing or understanding a single language [6, 34]. More recent work using word-embeddings to explore neologisms have demonstrated promise in both detecting and contextualising neologisms [37, 41]. It has also opened new avenues for conducting analyses on non-English corpora [32][6].

With regard to computational studies of the manosphere, corpus linguistic approaches have illuminated some specific characteristics of communities. For example, Heritage *et al* [18] examined more than 20 million terms used across the manosphere on Reddit, using keyword analysis, concordance line analysis and word frequencies, and qualitative analysis of keywords in their extended context. Their work specifically explored how women and girls are referred to in different groups on the manosphere. The authors found that conceptualisations of women were almost exclusively associated with physical attraction as the positive association and immorality and deception as negative associations. LaViolette and Hogan [23] analysed language use among a men's liberation community (/r/MensLib) and a men's rights community on reddit (/r/MensRights), combining natural language processing with discourse analysis to compare their different conceptualisations of some of the same issues of masculinity and feminism. The authors found that /r/MensLib embraced masculinity as an adjective, rather than an essential quality of manhood, and women as their peers. On /r/MensRights, being a man is discussed in essentialist terms, attributing blame to women for perceived discrimination and suffering. Previous work [11] identified hate and hostility as a feature of communities in the manosphere, categorising the type of misogyny prevalent in each group according to a set of codes derived from feminist literature. Ribreiro *et al* [31], using a larger sample of subreddits and conversations, were able to further characterise such groups as being more or less extreme in comparison with each other, and track participation through groups, demonstrating that newer, more extreme communities (such as MGTOW and incels) were "overshadowing" older communities.

Our approach incorporates three different knowledge sources (English dictionary, LDA model, word embeddings), as well as exploiting both semantic and subword information (via word embeddings). It can be oriented to specific topics (via seeding), and considers user usage of words. This is explained in detail in section 4. To our knowledge, none of the previous works look at such broad considerations.

3 CHARACTERISING THE MANOSPHERE

In this section, we look at characterisations about the manosphere that emerged from previous studies. Applying a grounded theory analysis [3], we collected examples of studies on the manosphere and compared findings across these examples, until we identified a set of major conflicts between self and external representations that appear again and again. We propose variations of these four conflicts as the 10 main themes (and four sub-themes) that will be used to analyse and study the subculture's jargon.

3.1 Self-Representations

Previous studies have explored how the manosphere represents itself.

3.1.1 Determinism and Essentialism. Communities in the manosphere, more generally, tend to have **deterministic views of masculinity and femininity, as well as the Laws of Attraction** [2, 30, 35]. More generally, the "red pill" construct, found across many communities in the manosphere[7], is an umbrella term for a set of beliefs that men and women are categorically different on the basis of

[6] https://lium.univ-lemans.fr/en/word-embeddings-temporels-neologismes-biais-de-genre-corpus-des-actualites-francaises/

[7] https://theredarchive.com/

a **perception of clearly defined sex**[8]. This deterministic worldview then extends to an idea of a natural or **evolutionary concept of attraction**, where women seek comfort and stability and men seek sexual fulfilment and fertility [9] (instead of such arrangements being the result of social stratification, for example [38]). According to the red-pill doctrine, men are disadvantaged by institutions like **marriage, the military and the labour force** [35], where they feel expected to continue to contribute despite a perceived **lack of return on their investment**. In some cases, this is believed to be a direct result of feminism [30].

3.1.2 Conflicts of Hegemony. Warren Farrell, a prominent personage in the men's rights movement argued that **men's understanding of power is flawed** and that, in the pursuit of economic or political power, men have lost sight over what it means to be in control over their lives [12]. In general, just the names of groups in the manosphere demonstrate a **conflict with what is perceived as hegemonic culture** and perceived exclusion (e.g. "involuntarily celibate" "going your own way") [26, 30].

3.1.3 Conflict of Masculinity. Unsurprisingly, within this larger conflict with mainstream society and expectations, there are many discussions within the manosphere about what it means to be a man or to be masculine. In several communities in the manosphere, the idea that there are 'alpha' and 'beta' males has given rise to a whole host of concepts (such as 'Chads', and 'Stacys')[10], the meaning of which has been explored in several studies referred to in section 2.2 [2, 15, 25]. Below we explore some of the themes that are more common in external representations of the manosphere.

3.2 External Representations

Studies of the manosphere demonstrate **ambiguous feelings toward fatherhood, absence from the family life and troubled relationships with women** as significant themes, as well as many different conflicts in relationship to **other men and their choices**, women and feminism [25]. We expand on these below.

3.2.1 Conflicts with Women and Feminism. Feminist characterisations of the manosphere focus on why certain ideas are tolerated or even appreciated [28]. For example, Gotell and Dutton's work explored the **contribution of men's rights activists to anti-feminist backlash**, in particular toward anti-rape movements [17]. Their work argued that **sexual violence is usurping father's rights as an emergent focus** in online spaces where men's rights activists converge. Analyses of responses to more recent campaigns, such as #metoo and #timesup, support this argument. From a feminist perspective, the misogynistic worldview of the manosphere stems from loss of power and perceived loss of societal role [19]. [42] argues that anger and frustration is directed toward women as a result of **men's own need to adapt to or cope with patriarchy**.

3.2.2 Conflicts with Men. As we have mentioned previously in section 2, several studies have explored the larger conflict within the manosphere about **what it means to be a man**, which men get access to which resources and why, and who are the losers and winners in this exchange [15, 16, 23].

[8]https://twitter.com/sciencevet2/status/1035246030500061184?lang=en
[9]https://theredarchive.com/post/210126
[10]https://bit.ly/3bxC8xz

3.3 Emergent Themes and Characterisations

From each of the analyses described above, we distilled several themes of interest to which jargon in the manosphere might relate. These themes are summarised in Table 1. **Opposition to feminism** and generally deterministic worldviews are two of the clearest themes of importance to both men in the manosphere and feminist scholars, who disagree with the essentialism of the manosphere. This worldview extends to **Laws of Attraction**, and the rules that such groups believe govern women's choices in relationships and marriage. Laws of attraction have subheadings because the laws are related in the literature to physique, race, class and other types of minority experiences. **Dehumanisation of women** is important to both men in the manosphere and feminist scholars, in the sense that the manosphere accepts certain language and hostility toward women, and this is believed to impact wider views on women in society. **Conflicts of masculinity** are found as well, though the conflict may be presented in different terms. Finally, there are categories are those that both feminists and men's rights activists acknowledge: there are real problems inside of these groups that need addressing, such as **mental health**, **personal suffering** and dissatisfaction with **family and relationships**. One additional category, which we have added to our analysis is **"trolling the manosphere"**. As members of the manosphere are not the only people communicating on the manosphere, "watchdog groups" and other individuals may be contributing to the development of jargon in these communities.

Main Thematic Categories	Abbr.	Examples
Conflict of Hegemony	CoH	normie, wagecuck, normtard
Conflict of Masculinity	CoM	soyboi, chadlite, betabux
Depravity	DEP	kissless, relationshit
Dehumanisation of women	DoW	femoid, awalt, roastie
General Determinism	DET	redpill, bluepill, blackpill
Dissatisfaction with Family/Relationships	FR	husbank, dependa, orbitor
Laws of Attraction	LOA	ricecel, looksmax, gymcel
Mental Health	MH	suifuel, lifefuel, suicel
Opposition to feminism	OF	feminazi, tradcon, cuck
Trolling Manosphere	TM	misogynatomy, rightcels
Subthemes in LOA		
Physique	LOA-PA	gyming, jawmog, nosecel
Race and racism	LOA-Race	Tyronne, ricecel, sandcel
Class and Status	LOA-Class	wagecel, statusmax, poorcel
Other Minorities	LOA-Other	neetcel, locationcel, incelqueer

Table 1: Examples of Categories of Significance

4 JARGON EXTRACTION APPROACH

The first research question tackles the automatic extraction of new words and jargon from user-generated conversations online. In this work we present a novel method that applies Latent Dirichlet allocation (LDA) modelling, as well as a word embedding model to semantically contextualise the extracted terms.

The proposed method has a pipeline of several steps: (i) seeding, (ii) frequency filtering, (iii) dictionary-based filtering, (iv) topic

checking, (v) embedding expansion, (vi) embedding filtering, and (vii) user usage filtering. Each of these steps implements a set of operations that either reduces or expands the set of candidate terms, and each step's output is used as input for the following. Parameters of the extraction method can be changed, and are used to configure the different steps. When finished, the method outputs a list of jargon candidates. The proposed steps are described next.

Firstly, the **seeding** step can be set to initialise the candidate extraction, in order to centre around a set of interesting terms for the task at hand, which could be included in a domain lexicon. For example, if we were only interested in extracting jargon terms related to hate, we could use a lexicon of hate, such as the one provided by [11], as seeding. Alternatively, the method could be initialised without seeding (i.e., by considering the full vocabulary of the provided input corpora).

In the second step, the initial set is **filtered by frequency**. All terms that appear in the input corpora less than a certain threshold are removed. This filtering intends to filter out terms that are not commonly used by the subculture (e.g., misspelling of known words).

Following the frequency filtering, a **dictionary-based filtering** is performed. Since the aim of this work is to discover specialised emergent vocabulary (i.e., new distinct terms used by the subculture) a dictionary (in this case an English dictionary) is used to filter all terms that appear in the mentioned dictionary. Thus, we ensure that the terms that remain after this step do not belong to an English dictionary.

Once the first steps of the proposed method are completed, we perform a **topic checking**. The purpose of this step is: (i) to identify which topics (from the ones that are extracted using the full input corpora) have been lost during the previous filtering steps and, (ii) to find associations between jargon terms and specific topics. To do so, we train an LDA model (previous to the execution of the pipeline) and obtain the topics that emerge from the input corpora. To obtain the optimal number of topics the LDA model is optimised attending to the coherence, as done in [33]. The output of this step is a series of topics, each of them represented as a bag of terms. We then match each of the candidate jargon terms (obtained after the **dictionary-based filtering**) to the bag of words of each topic and calculate those topics that are not covered by any jargon term (i.e., the themes of discussion that are not covered by the subsculture's specialised vocabulary) as as well as the different topics to which each jargon term belongs. The next step, **embedding expansion** performs an expansion of the set of candidate jargon terms by means of embedding similarity. For each of the candidate jargon terms we find terms that are close in the embeddings space (considering a certain threshold). This serves two purposes: (i) to contextualise jargon terms by providing for them a set of terms that are morphologically and semantically close in the embedding space and (ii) to find new jargon terms (which is particularly important when seeding is performed, since it allows us to expand the initial set of seeds and to find additional jargon terms). Note that during the embedding expansion, both jargon and not jargon terms may be extracted. To differentiate them, terms are marked, indicating if they appear or not in the dictionary. As done with **topic checking**, this method also requires the training of a word embedding model prior to the execution of the pipeline [29]. For this work, we selected the recent word embedding model presented in [5], which considers both semantic and subword similarity.

In a posterior step, the method performs the **embedding filtering**. The selected word embedding model encodes both semantic and morphological similarity. Exploiting this feature, the method removes words that are morphologically similar, or similar subword-wise. These are mainly variations of a jargon word. The method retains the most common of these variations and removes the rest. The filtering of such variations is checked using the Levenshtein distance.[11] to ensure that only morphological variations are eliminated. Two thresholds are used in this step: one for the embedding filtering, and one for the Levenshtein distance.

Finally, the method concludes with the user **usage filtering**. This steps removes the terms that are used by fewer than a certain number of distinct users. By selecting this user usage parameter, the method assures that the jargon terms included in the final candidate list are adopted by a wide range of users in the subculture.

5 ANALYSIS SET-UP

In this section we describe the different online communities selected for this study as well as the analysis conducted.

5.1 Selection of Online Communities

Based on a previous study [11] seven Reddit communities have been selected that are different representatives of the Manosphere's subculture. This list is not exhaustive. Nor does it include groups that are too broad in their focus (such as r/theredpill, or r/mensrights).

- *r/MGTOW*: this is a subreddit of 'men going their own way', in which men claim that they wish to simply live a life without the interference from women.
- *r/Braincels*: this is the main incel subreddit since *r/incels* was removed from Reddit in November 2017 for violating site-wide rules. Some members of Braincels were also self-reported members of the website incels.me (now defunct), and the more current incels.is[12] or similar non-Reddit websites, where more violent content is posted.
- *r/Trufemcels*: this is a subreddit of women who are self-described incels. Male incels occasionally remark that it is not possible for a female incel to exist, given the advantages of women over men in finding a sexual partner.
- *r/IncelsWithoutHate*: this is a subreddit of individuals who are self-described as both incel but non-violent.
- *r/Inceltears*: this is a subreddit dedicated to calling out Incels. They screenshot and post particularly egregious content from r/braincels, incels.me, incels.is and other incel communities. They are partly responsible for a large number of incel communities being closed down.
- *r/IncelsInAction*: this is a subreddit that monitors activity from other incel communities, similarly to *r/Inceltears*.
- *r/badwomensanatomy*: this is a subreddit focusing on women's bodies in a misogynistic way. This group shares some of the same misogynistic vocabulary, but not the same ideology as other groups in this set.

[11] https://en.wikipedia.org/wiki/Levenshtein_distance
[12] https://incels.is/

Data from the above seven communities has been gathered from their inception until January 2019 via the pushshift API[13]. This has led to the collection of 301,078 conversations and a total of 5,674,303 comments in those conversations (see Table 2). Table 2 shows a summary of the collected data, including the online community, its number of posts, and the dates of the first and last post.

5.2 Jargon Extraction

In this section we specify the parameters used during the jargon extraction methodology. We start with a corpus of nearly 6M posts. We use this corpus to extract an LDA model and a word embedding model previous to the execution of the pipeline. The vocabulary size of this corpus is 1,138,898 terms. The number of LDA topics extracted in this process is 60.

We start the jargon extraction pipeline by considering no seeding, since we are interested on extracting all possible jargon terms. That means that our pipeline departs from the 1,138,898 terms from the vocabulary. After frequency filtering and removing punctuation, using a frequency threshold of 20, we remain with 51,588 terms. We discard then all of the terms appearing in an English dictionary remaining with 3,426 candidate jargon terms. We perform the embedding expansion to contextualise these terms using an embedding expansion factor value of 5.244 and considering 30 as the maximum number of neighbours to include in the expansion. Note that, since we start the pipeline with the complete vocabulary (no seeding), no additional jargon terms have been added to the list of candidates after this step. However, the 3,426 candidate terms are now contextualised by means of a set of 17,968 additional terms (that are not jargon). The 3,426 candidate jargon terms are then filtered using the embedding filtering process (using as similarity threshold 0.85 and as Levenshtein distance threshold 3), remaining with a set of 2,865 terms that are further filtered in the final step (usage filtering) to 2,615 jargon terms. We set the number of users with which the filtering is done to 10.

5.3 Jargon Annotation

Automatically identified terms through the methodology described in Sect. 4 were manually annotated according to the coding scheme of major themes and sub-themes as described in 3.3 and Table 1.

One annotator with deep knowledge of the manosphere and familiarity with the Reddit platform validated each term manually, in context, on Reddit. In addition, the terms were reviewed on the Urban Dictionary[14] and on the Web. Each term was then assigned to one of the major themes and sub-themes as illustrated in Table 1. During the process of validating terms, however, the annotator developed several other codes to denote different types of language that would be excluded from our study of jargon. These are described in Table 3 and include common-use terms across Reddit, real people and places, as well as other user groups or platform-based jargon. 265 of the terms were not annotated because they were long strings of letters or initialisms that were not possible to code or because the context was not sufficient to categorise them (e.g., whitecel, westcoastincel).

[13]https://pushshift.io/
[14]https://www.urbandictionary.com/

Community	numPosts	minDate	MaxDate
MGTOW (MG)	168124	2011-06-04	2019-01-11
badwomensanatomy (BwA)	13010	2014-01-02	2019-01-11
IncelsWithoutHate (IwH)	2309	2017-04-09	2019-01-11
IncelTears (IT)	15679	2017-05-19	2019-01-11
IncelsInAction (IiA)	330	2017-06-24	2019-01-10
Braincels (BC)	96545	2017-10-21	2018-10-01
Trufemcels (TF)	5081	2018-04-04	2019-01-11

Table 2: Summary of Reddit collected data.

Main Exclusion Categories	No.	Abbr.	Examples
Common Usage	596	CU	whorish, wimmin, terfy
Clear Spelling Mistakes	352	SP	because, alow
Real People or Characters	63	RP	weinstein, Peterson
Real Words or Things	237	RW	guidestones, hive mind
Reddit Terminology and Conventions	72	DoW	upbeat, brigading, unsubbing
Users or Groups on Reddit	298	UG	asablackman, bad anatomy
Not Annotated	265	NA	westcoastincel, autoblow, ammm

Table 3: Exclusion Categories

6 ANALYSIS RESULTS

In this section we display the results of our analyses and discuss key insights.

6.1 Extracted Jargon

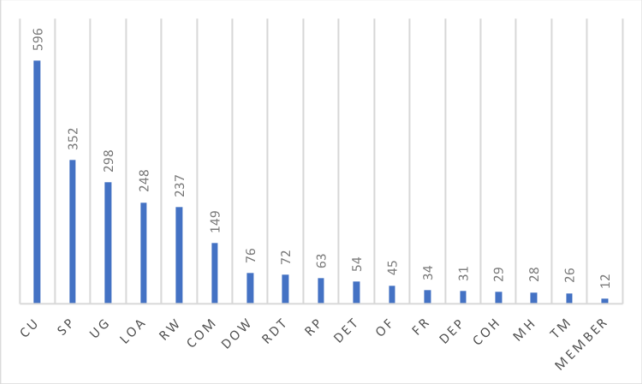

Figure 1: Jargon terms per category

Based on our methodological pipeline we automatically identified and manually annotated 2,615 candidate jargon terms in different categories. Figure 1 shows the distribution of these annotations per category. Terms that fall into the Exclusion Categories, identified during the annotation and validation process of the extracted vocabulary (see Table 3), are discarded, since they do not represent the subculture's jargon. Terms that fall into the Main Thematic Categories (see Table 1), 732 terms in total, are jargon terms specific from the Manosphere.

6.2 Global Analysis of Jargon

We can observe (Figure 1) that the most prominent thematic categories categorising this jargon are: Laws of Attraction, followed by Conflict of Masculity, Dehumanisation of Women, Opposition to Feminism and Dissatisfaction with Family and Relations (all these categories presenting 30 or more jargon terms). However, while we observe this distribution of jargon per categories globally, the distribution per individual communities (or Reddit Forums) differ, in some cases significantly, from the global distribution (see Figure 2). Hence, in terms of defining a subculture through shared language, the manosphere looks more like a subsociety, in which there are some shared terms and experiences, but also many other that are not shared, or for which there are unique perspectives [13]. In the next section we analyse the distribution of jargon per community, deriving observations and comparison across them.

6.3 Comparative Analysis of Jargon

Thematic	IT	IiA	TF	BC	MG	IwH	BWA
CoH	16	2	7	29	18	13	2
CoM	60	4	34	138	128	43	5
CU	256	10	161	501	551	140	95
DEP	19	2	18	29	29	18	3
DET	36	7	33	51	46	30	12
DoW	33	0	26	64	59	21	13
FR	5	0	5	21	34	7	2
LOA	130	11	97	245	85	99	5
MH	17	2	17	27	16	18	1
OF	12	0	6	30	42	9	3
TM	13	0	2	23	12	8	7
TOTAL	597	38	406	1158	1020	406	148
Exclusion							
Member	8	3	6	9	12	7	3
RDT	49	5	33	68	66	23	27
RP	21	1	6	45	53	12	4
RW	85	3	64	171	203	55	51
SP	202	9	125	313	340	117	58
UG	147	9	73	224	223	76	93
TOTAL	512	30	307	830	897	290	236

Table 4: Number of Jargon terms per category used in each community.

Table 4 shows the distribution of terms per community including both, the Main Thematic Categories (top part of the table) and the Exclusion Categories (bottom part of the table). To complement this table, Figure 2 focuses on the distribution of the 732 identified jargon terms per community and Main Thematic Categories. The figure displays the percentage of usage of each category within each community, which helps displaying a better comparative.

6.3.1 Prominent Themes. We can observe from this comparative table how Laws of Attraction is a very prominent category for all groups except for MGTOW and badwomensanatomy. Within this category, most of the communities also focus on Physique as prominent subcategory. For MGTOW, the most prominent category is Conflict of Masculinity. The most common category for badwomensanatomy, our control group, has a high number of terms associated Dehumanisation of Women, but this group also has the higher percentage of jargon for Trolling the Manosphere among other groups (e.g. biotruth, biotroof[15]). Compared to the other groups, "watchdog group" IncelsInAction, does not appear to be fully embedded in the jargon with the other groups, as it has no very prominent themes.

6.3.2 Shared Jargon. An interesting observation is how jargon is shared across the different subgroups. Figure 3 displays this analysis. The diagonal contains the number of jargon terms (out of the 732 identified ones) used by each group. The rest of the cells display the number of jargon terms shared between two given groups (independently on whether other groups also use those terms). As we can see all communities share most of their jargon with Braincels, which is indeed the most prominent community on the use of jargon, using 657 of the identified jargon terms. This community shares most of its jargon with MGTOW, followed by IncelTears and IncelsWithoutHate. As Braincels is believed to have been started with original members of the banned subreddit r/incels[16], it may have a diverse and potentially frustrated base.

Truefemcels also shares vocabulary with Braincels. This is interesting, because direct, hostile misogyny is not as evident in their discussions on Reddit [11]. Looking more deeply at their shared jargon terms in more detail qualitatively, these are largely about physical attractiveness and conflicts with mainstream expectations of physical beauty (e.g., normies, stacy, chadlite). Truefemcels also shares a higher representation of the category LOA Race. With MGTOW, Truefemcels appears to share *their* central crises of hegemony and masculinity. For example, the two groups share several terms that are denominalisations and adjectivisations of "Chad", the all-American male[17] (e.g., chadding, chaddier, chaddest).

IncelswithoutHate and Trufemcels, our two non-violent incel groups share half of their specialised vocabulary from Figure 3. Deterministic world views and issues with physical attraction appear to be the words they most have in common, as well as several words that are related to coping mechanisms (e.g., looksmaxxing, looksmatch, gymcope), which is what we would expect given our understanding of the basic ideas of involuntary celibacy.

6.3.3 MGTOW vs. Braincels. Since MGTOW and Braincels share the most terms among the different communities we have done an analysis on the jargon terms specifically shared by these communities (that do not appear in any of the other analysed groups). A breakdown of how these terms fit into our major thematic categories can be see in Figure 4. What divides these groups ideologically, according to socio-linguistic studies, is that MGTOW claim that they do not want the company of women and incels (of which Braincels are considered part) appear to want relationships with women (by virtue of the involuntary part of their celibacy). Terms that express Conflict with Masculinity and Dehumanisation of Women, however, are commonly shared between the two groups.

6.4 Analysis of Individual Communities and Jargon

In this subsection, we provide more in-depth analysis of jargon that we felt showed interesting insights.

[15] words that make fun of red pill, deterministic explanations of behaviour
[16] https://www.theguardian.com/technology/2017/nov/08/reddit-incel-involuntary-celibate-men-ban
[17] https://en.wikipedia.org/wiki/Chad_(slang)

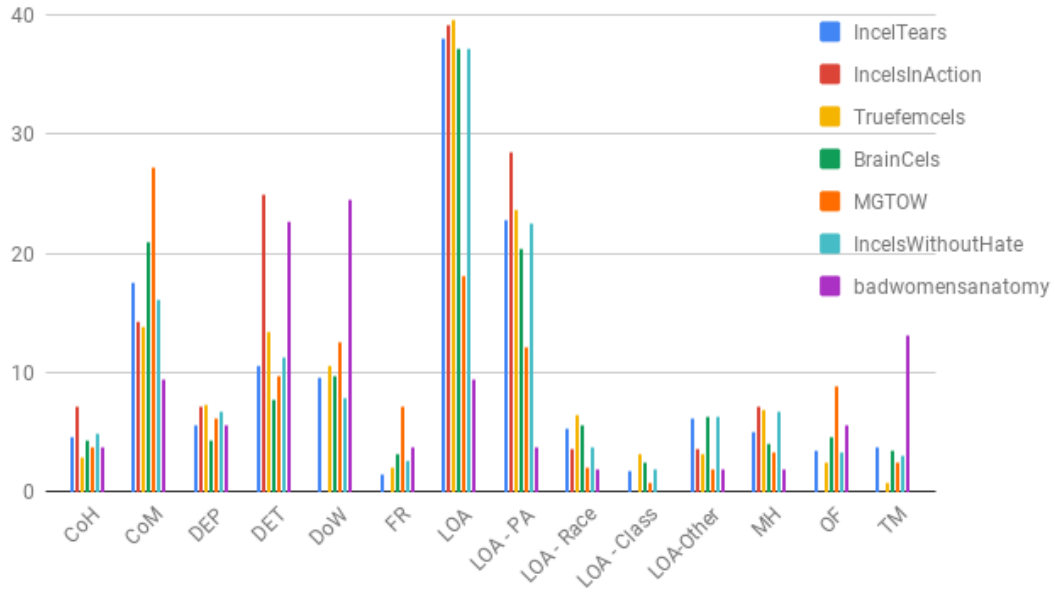

Figure 2: Percentage of jargon terms in each category per community

	IT	IiA	TF	BC	MG	IwH	BwA
IncelTears	341	26	172	334	237	196	39
IncelsInAction	26	28	24	28	25	23	3
Truefemcels	172	24	245	242	184	153	30
BrainCels	334	28	242	657	418	260	44
MGTOW	237	25	184	418	469	196	42
IncelsWithoutHate	196	23	153	260	196	266	25
badwomensanatomy	39	3	30	44	42	25	53

Figure 3: Number of Jargon terms shared between communities

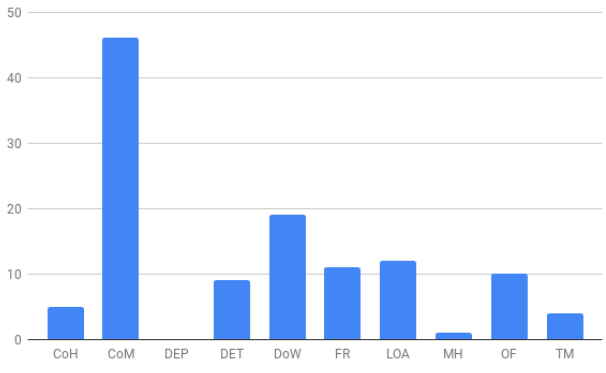

Figure 4: Shared Terms for MGTOW and Braincels by Frequency of Associated Theme

6.4.1 Braincels. Braincels has many terms that were coded into the category of physical attraction, like other incel groups. When looking at the sub-themes, however, Braincels tended to focus on physical things that one cannot change (e.g., baldingcel, babyfacecel), or systemic issues such as poverty (e.g., wagecel, poorcel) and racism/xenophobia (e.g., beancel, browncels). Where other incel groups speak about maximising chances with women (words ending with the suffix -maxx), Braincels appears to have more need for terms that express the perceived futility of wanting a mate or of trying to be successful in general. This more resembles MGTOW, in that the desire to find a mate is overcome by perceived systemic and social challenges.

6.4.2 MGTOW. MGTOW uses the highest percentage of vocabulary from family and relationships (FR) across groups. Words like "divorcerape" and "pussypass" are surfaced in the analysis. This fits with previous self- and external representations of MGTOW that dissatisfaction with family and relationships are a driving motivation to take part, and that hostile expression of this is characteristic of this community [11, 31]. Where these ideas are surfacing, between and among these communities, would be a good subject of future research.

6.4.3 "Watchdog" Groups. Finally, one "watchdog group" (IncelTears) does appear to be aware of new terms across all categories. They share the greatest number of new terms with MGTOW and Braincels (who have been identified in previous studies as the most hostile incel-related groups in previous work) [31].

6.5 Contextualisation based on word-embeddings

One of the research questions posed in this work is whether contextualisation based on word-embeddings could help us to better

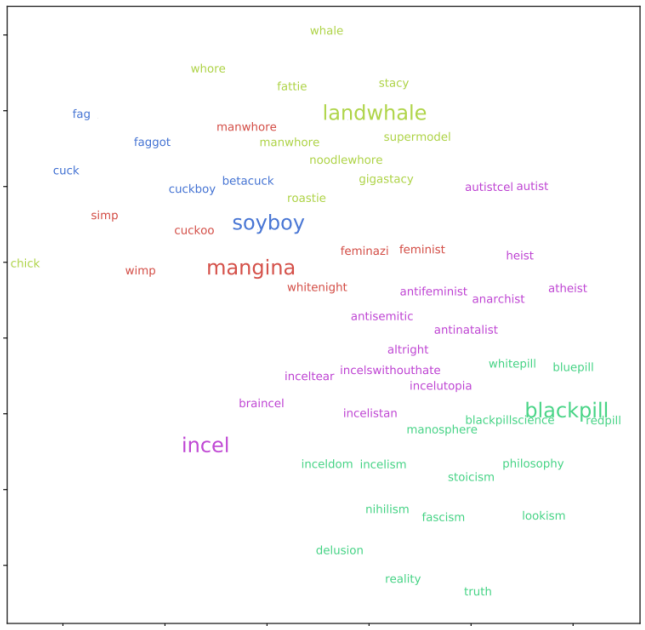

Figure 5: Examples of characterisation via word-embeddings

understand the meaning of jargon. An example of such contextualisation for five jargon terms (landwhale, soyboy, mangina, blackpill and incel) is displayed in Figure 5. The contextualisation of each term is display in a different colour. We define as context the set of terms that are closed to the jargon term in the embedding space and that can help us to better understand its meaning. For example, if one is not sure of the meaning of landwhale, one can see that it falls into the same category as both model and Stacy[18], as well as roastie and fattie (both of which are negative terms). From there, one can associate the term Landwhale with perceptions of a woman's physical attractiveness.

As a way to assess whether word-embeddings do indeed facilitate the understanding of jargon, the results of the human annotation with the Main Thematic Categories have been compared against the generated word-embeddings by calculating the distance within the embedding space. To do so, we have first computed a baseline distance. This is the embedding distance between a number of randomly selected words. The number of terms used for computing the baseline was 120, which is the average number of terms for all categories. We have then computed the distance per category and compare such distance with the baseline. The spatial metric was the cosine distance. Results of this computation are displayed in Table 5. The lower the result, the lower the distance of the words of the category in the embedding space.

Our results show that the word groupings identified by the human annotator do indeed align with the grouping of the terms in the embedding space. This is true for all categories except for DEP (Depravity). It is important to notice that DEP includes terms that may denote different aspects of suffering or different types

[18]Stacys are attractive females, usually white [30]

Table 5: Distance per category

Category	Distance
Baseline	0.615
LOA	0.486
CoM	0.490
DoW	0.528
DET	0.423
OF	0.512
FR	0.585
DEP	**0.631**
CoH	0.502
MH	0.524
TM	0.502

of suffering that may be semantically dissimilar. Overall, however, we can say that close terms in the embedding space may provide a semantic context and hence, a support for understanding jargon.

7 DISCUSSION

We proposed a combination of computational and socio-linguistic methods to automatically identify and analyse jargon from large amounts of data. This enabled us to study a subsociety, such as the manosphere, by means of the extraction and understanding of its specialised vocabulary. Our proposed method advances the state-of-the-art by: (i) incorporating different knowledge sources in the identification of jargon (semantic information, subword information, topic seeding and user usage of words), (ii) exploiting word embeddings for expansion, filtering and contextualisation and, (iii) providing an in-depth analysis and empirical validation of the identified jargon by means of its alignment with key thematic categories of the subculture (identified from previous literature).

We observed that the studied groups create new terms and language for experiences not widely shared across the manosphere. Not all groups prioritise being attractive for women, nor do they all develop specialist vocabulary around negative experiences with women. Not all groups have the same conflict of masculinity. For some, it is purely physical and for others, it is economic and social. In a way, what seems to emerge is a picture of the manosphere engaging in a discussion about intersectionality [9], in which men are positioning themselves against one another.

The increasingly hostile, misogynistic alt-right tone of some parts of the manosphere is still concerning [31]. However, the manosphere is not only full of white, angry men. There are also men of colour, struggling with systemic racism that extends to beauty ideals and status. However, when they blame women for their problems, this misogyny is coupled with racism, compounding the impacts for women of colour. This issue needs future attention[19,20]

One limitation of our proposed method is that it is focused on identifying new terms, rather than identifying new meaning to existing terms. Hence, more in-depth studies could reveal deeper insights. However, in comparison with smaller scale socio-linguistic studies, computational studies of this kind create inroads for looking at data historically and at scale. What we have uncovered in this

[19]https://melmagazine.com/en-us/story/currycels-and-the-unsurprising-racism-of-the-incel-community
[20]http://blackyouthproject.com/its-not-just-white-incels-we-need-to-talk-about-the-black-manosphere-too/

research can now be used to examine shifts and emergence over time, with annotations enriched by socio-linguistic studies.

The automatically extracted terms were categorised by one annotator. While it would be desirable to have more annotators from which agreements/disagreements could be observed, categorising jargon requires in-depth knowledge of the subculture. Annotators with this expertise are scarce, hence we have opted for a categorisation of the jargon conducted by one expert annotator rather than by a wider set of annotators that may not have sufficient knowledge.

It is also important to notice that our extraction methodology captured many jargon terms but also noise, due to the platform and the nature of conversations (with many Reddit terms and mentions of unique user or group names). We also identified other borrowed specialised vocabulary or slang (which we categorised as common-use terms -CU). Looking into these exclusion categories provided a richer qualitative analysis in terms of the people, user groups, common use terms, and regular words that appear across communities.

The exploration of the alignment between the topics extracted during our analysis pipeline and the topic areas that we identified in the literature is part of our future research. In a similar way, a in-depth study on the temporal evolution of the detected jargon, as well as its contextualisation in reference to non-misogynistic cultures constitute research areas that will be addressed in the future. Additionally, new data could be captured, including further subreddits and other internet forums.

8 CONCLUSIONS

In this paper, we proposed a combination of computational and socio-linguistic methods to extract and contextualise novel jargon terms in the manosphere. Exploring specialist vocabulary that emerged from this community, we were able to compare the potential need for different jargon, as evidenced by innovation of new words. We explored what purpose it might serve and how it relates to different lived experiences. In turn this has made it possible to differentiate communities on the manosphere computationally.

REFERENCES

[1] [n.d.]. Alek Minassian. https://www.telegraph.co.uk/news/2018/04/24/do-know-alek-minassian-arrested-toronto-van-attack/. Accessed: 2019-01-27.

[2] [n.d.]. INCELS. http://haenfler.sites.grinnell.edu/subcultures-and-scenes/incels/#ffs-tabbed-311. Accessed: 2019-01-27.

[3] Mohammed Ibrahim Alhojailan. 2012. Thematic analysis: A critical review of its process and evaluation. *West East Journal of Social Sciences* 1, 1 (2012), 39–47.

[4] Stephane J Baele, Lewys Brace, and Travis G Coan. 2019. From "Incel" to "Saint": Analyzing the violent worldview behind the 2018 Toronto attack. *Terrorism and Political Violence* (2019), 1–25.

[5] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics* 5 (2017), 135–146.

[6] Emmanuel Cartier. 2017. Neoveille, a web platform for neologism tracking. In *Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. 95–98.

[7] Elaine Chaika. 1980. Jargons and language change. *Anthropological Linguistics* 22, 2 (1980), 77–96.

[8] David Claerbaut. 1972. Black Jargon in White America. (1972).

[9] Patricia Hill Collins and Sirma Bilge. 2016. *Intersectionality*. John Wiley & Sons.

[10] C Paul Cook. 2010. *Exploiting linguistic knowledge to infer properties of neologisms*. University of Toronto Toronto, Canada.

[11] Tracie Farrell, Miriam Fernandez, Jakub Novotny, and Harith Alani. 2019. Exploring misogyny across the manosphere in reddit. In *Proceedings of the 10th ACM Conference on Web Science*. 87–96.

[12] Warren Farrell. 1996. *The myth of male power*. Berkeley Publishing Group.

[13] Gary Alan Fine and Sherryl Kleinman. 1979. Rethinking subculture: An interactionist analysis. *American journal of sociology* 85, 1 (1979), 1–20.

[14] Michael Flood. 2004. Backlash: Angry men's movements. In *The battle and backlash rage on: Why feminism cannot be obsolete*. Xlibris Press, 261–278.

[15] Debbie Ging. 2017. Alphas, betas, and incels: Theorizing the masculinities of the manosphere. *Men and Masculinities* (2017), 1097184X17706401.

[16] Debbie Ging and Eugenia Siapera. 2018. Special issue on online misogyny.

[17] Lise Gotell and Emily Dutton. 2016. Sexual violence in the 'manosphere': Antifeminist men's rights discourses on rape. *Int. J. Crime, Justice and Social Democracy* 5, 2 (2016), 65–80.

[18] Frazer Heritage, Veronika Koller, Alexandra Krendel, and Abi Hawtin. 2019. MANTRaP: A Corpus Approach to Researching Gender in Online Misogynist Communities. *12th BAAL LGaS SIG* (2019).

[19] Christa Hodapp. 2017. *Men's rights, gender, and social media*. Rowman & Littlefield.

[20] Emma A Jane. 2016. *Misogyny online: A short (and brutish) history*. Sage.

[21] Christopher M Johnson. 2001. A survey of current research on online communities of practice. *The internet and higher education* 4, 1 (2001), 45–60.

[22] Peter Langman. 2014. Elliot Rodger: An Analysis. *The Journal of Campus Behavioral Intervention* 2 (2014), 5–19.

[23] Jack LaViolette and Bernie Hogan. 2019. Using platform signals for distinguishing discourses: The case of men's rights and men's liberation on reddit. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13. 323–334.

[24] Lauraine Leblanc. 1999. *Pretty in punk: Girls' gender resistance in a boys' subculture*. Rutgers University Press.

[25] Mary Lilly. 2016. *'The World is Not a Safe Place for Men': The Representational Politics of the Manosphere*. Ph.D. Dissertation. Université d'Ottawa/University of Ottawa.

[26] Alan MacLeod. 2019. Book review: Kill all normies: Online culture wars from 4chan and Tumblr to Trump and the alt-right.

[27] Kathleen M MacQueen, Eleanor McLellan, David S Metzger, Susan Kegeles, Ronald P Strauss, Roseanne Scotti, Lynn Blanchard, and Robert T Trotter. 2001. What is community? An evidence-based definition for participatory public health. *American journal of public health* 91, 12 (2001), 1929–1938.

[28] Alice E Marwick and Robyn Caplan. 2018. Drinking male tears: language, the manosphere, and networked harassment. *Feminist Media Studies* 18, 4 (2018).

[29] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).

[30] Angela Nagle. 2017. *Kill all normies: Online culture wars from 4chan and Tumblr to Trump and the alt-right*. John Hunt Publishing.

[31] Manoel Horta Ribeiro, Jeremy Blackburn, Barry Bradlyn, Emiliano De Cristofaro, Gianluca Stringhini, Summer Long, Stephanie Greenberg, and Savvas Zannettou. 2020. From Pick-Up Artists to Incels: A Data-Driven Sketch of the Manosphere. *arXiv preprint arXiv:2001.07600* (2020).

[32] Andrés Torres Rivera and Juan-Manuel Torres-Moreno. 2019. Detecting new word meanings: a comparison of word embedding models in Spanish.. In *CORIA*.

[33] Michael Röder, Andreas Both, and Alexander Hinneburg. 2015. Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining*. 399–408.

[34] O'Donovan Ruth and O'Neill Mary. 2008. A systematic approach to the selection of neologisms for inclusion in a large monolingual dictionary. In *Proceedings of the 13th Euralex International Congress*. 571–579.

[35] Rachel M Schmitz and Emily Kazyak. 2016. Masculinities in cyberspace: An analysis of portrayals of manhood in men's rights activist websites. *Social Sciences* 5, 2 (2016), 18.

[36] Suzanne M Shultz. 1996. Medical jargon: ethnography of language in a hospital library. *Medical reference services quarterly* 15, 3 (1996), 41–47.

[37] Brooke A Taylor, Brendan Tracey, and Vanessa Ferdinand. [n.d.]. Measuring the Change in Semantic and Sentiment of Neologisms Against An Axis. ([n. d.]).

[38] Pierre L Van Den Berghe. 1960. Hypergamy, hypergenation, and miscegenation. *Human Relations* 13, 1 (1960), 83–91.

[39] Tony Veale and Cristina Butnariu. 2010. Harvesting and understanding on-line neologisms. *Cognitive Perspectives on Word Formation* 221 (2010), 399.

[40] Quirin Würschinger, Mohammad Fazleh Elahi, Desislava Zhekova, and Hans-Jörg Schmid. 2016. Using the Web and Social Media as Corpora for Monitoring the Spread of Neologisms. The case of 'rapefugee','rapeugee', and 'rapugee'.. In *Proceedings of the 10th Web as Corpus Workshop*. 35–43.

[41] Nasser Zalmout, Kapil Thadani, and Aasish Pappu. 2019. Unsupervised Neologism Normalization Using Embedding Space Mapping. In *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*. 425–430.

[42] Donna Zuckerberg. 2018. *Not All Dead White Men: Classics and Misogyny in the Digital Age*. Harvard University Press.